

THE CONSTRUCTION OF FINITE DIFFERENCE APPROXIMATIONS TO ORDINARY DIFFERENTIAL EQUATIONS*

EUSEBIUS J. DOEDEL†

Abstract. Finite difference approximations of the form $\sum_{i=-r_j}^{s_j} d_{j,i} u_{j+i} = \sum_{i=1}^{m_j} e_{j,i} f(z_{j,i})$ for the numerical solution of linear n th order ordinary differential equations are analyzed. The order of these approximations is shown to be at least $r_j + s_j + m_j - n$, and higher for certain special choices of the points $z_{j,i}$. Similar approximations to initial or boundary conditions are also considered and the stability of the resulting schemes is investigated.

1. Introduction. We consider a method for constructing finite difference approximations to the linear ordinary differential equation

$$(1.1) \quad Ly(x) \equiv y^{(n)}(x) + \sum_{k=0}^{n-1} a^k(x) y^{(k)}(x) = f(x), \quad 0 \leq x \leq 1.$$

Approximations to boundary conditions are also considered. The consistency and stability properties of these approximations are investigated and some numerical results are given.

Numerous texts treat finite difference approximations to differential equations, for example the books by Collatz (1966), Keller (1968), Gear (1971) and Keller (1976). The general theory of difference methods for boundary value problems in ordinary differential equations is contained in papers by Grigorieff (1970) and Kreiss (1972). The construction of difference methods has been considered also by for example Birkhoff and Gulati (1974), Swartz (1974), Doedel (1976), Lynch and Rice (1976), Keller and Pereyra (1976) and many more. In particular the results reported by Lynch and Rice (1976) are essentially the same as those given in the present paper. The main difference between the two studies, which were carried out independently, is the fact that in this paper we also include noncompact approximations. High order compact difference approximations have also been investigated by Osborne (1964), (1967), (1974). The underlying ideas in his work also closely resemble those in the present paper, but the techniques employed for analyzing and implementing the method are somewhat different.

The extension of the method to the case of linear systems is essentially a matter of notation. Somewhat more work is required for nonlinear equations. This is investigated by the author in a separate paper (1977). The case of partial differential equations is definitely more involved, but the author is convinced that such a study would lead to many useful results. A preliminary report on extensions to elliptic boundary value problems in 2 and 3 dimensions has already appeared. (See Lynch and Rice (1975).)

2. The method. Define a mesh $S_h \equiv \{x_i : x_0 < x_1 < \cdots < x_J\}$. Let $h_j \equiv x_j - x_{j-1}$ ($1 \leq j \leq J$), and $h \equiv \max_j h_j$. We assume that $\min_j h_j \geq c_1 h$ for some positive constant c_1 that is independent of h . For any function $w(x)$ defined on S_h let $w_j \equiv w(x_j)$.

We consider finite difference approximations to the differential equation (1.1) at the point $x = x_j$ of the form

$$(2.1) \quad L_h u_j \equiv \sum_{i=-r_j}^{s_j} d_{j,i} u_{j+i} = \sum_{i=1}^{m_j} e_{j,i} f(z_{j,i}).$$

* Received by the editors September 1, 1976, and in revised form March 21, 1977. This work was supported by the National Research Council of Canada under a Postgraduate Scholarship and by the U.S. Energy Research and Development Administration under Contract AT-04-3-767, Project Agreement 12.

† Applied Mathematics, California Institute of Technology, Pasadena, California. Now at Mathematics Department, Vanderbilt University, Nashville, Tennessee 37235.

To simplify the notation the subscript j is frequently omitted, so that for example $z_{j,i}$ becomes z_i . The points z_i in (2.1) need not coincide with the mesh points, but they are required to satisfy $|z_i - x_j| \leq c_2 h$ and $|z_{i_1} - z_{i_2}| \geq c_3 h$ if $i_1 \neq i_2$. Here c_2 and c_3 are positive constants.

Let \mathbb{P}_d denote the space of all polynomials with degree not exceeding d . Let a basis of $\mathbb{P}_{r+s+m-1}$ be given by polynomials $\{w_k(x)\}_{k=-r}^s \cup \{v_k(x)\}_{k=1}^{m-1}$ that satisfy

$$(2.2a) \quad w_k(x_{j+l}) = \delta_{k,l}, \quad -r \leq k \leq s, \quad -r \leq l \leq s,$$

and

$$(2.2b) \quad v_k(x_{j+l}) = 0, \quad 1 \leq k \leq m-1, \quad -r \leq l \leq s.$$

Here $\delta_{k,l}$ is the Kronecker delta function. For example a basis satisfying (2.2a, b) is given by

$$(2.3a) \quad w_k(x) = \prod_{\substack{l=-r \\ l \neq k}}^s \frac{(x - x_{j+l})}{(x_{j+k} - x_{j+l})}, \quad -r \leq k \leq s,$$

and

$$(2.3b) \quad v_k(x) = h^{r+s+k} (x - x_j)^{k-1} \prod_{l=-r}^s (x - x_{j+l}), \quad 1 \leq k \leq m-1.$$

The first $r+s+1$ basis polynomials given in (2.3a) are the Lagrangian interpolating coefficients with respect to the meshpoints x_{j+i} , $(-r \leq i \leq s)$.

The coefficients d_i and e_i in the difference equation (2.1) can now be determined by requiring that (2.1) is satisfied for all $p \in \mathbb{P}_{r+s+m-1}$, i.e.

$$(2.4a) \quad \sum_{i=-r}^s d_i p(x_{j+i}) = \sum_{i=1}^m e_i Lp(z_i),$$

where for $p(x)$ we take $w_k(x)$ $(-r \leq k \leq s)$, and $v_k(x)$ $(1 \leq k \leq m-1)$, consecutively. For uniqueness it is necessary to impose some normalization condition such as

$$(2.4b) \quad \sum_{i=1}^m e_i = 1.$$

The equations (2.4a, b) lead to a linear system of the form

$$(2.5) \quad \begin{pmatrix} I & -\mathcal{A} \\ O & -\mathcal{B} \\ O^T & \gamma^T \end{pmatrix} \begin{pmatrix} d \\ e \end{pmatrix} = \begin{pmatrix} o \\ o \\ 1 \end{pmatrix},$$

where $d \equiv (d_{-r}, \dots, d_s)^T \in \mathbb{R}^{r+s+1}$, $e \equiv (e_1, \dots, e_m)^T \in \mathbb{R}^m$, $\gamma^T \equiv (1, 1, \dots, 1)^T \in \mathbb{R}^m$ and o denotes a zero vector of appropriate dimension. $I \in \mathcal{M}_{r+s+1, r+s+1}$ is the identity and $O \in \mathcal{M}_{m-1, r+s+1}$ the zero matrix. Here $\mathcal{M}_{k,l}$ is the space of matrices with k rows and l columns. Further, the elements of $\mathcal{A} \in \mathcal{M}_{r+s+1, m}$ and $\mathcal{B} \in \mathcal{M}_{m-1, m}$ are given by

$$(2.6) \quad [\mathcal{A}]_{k,l} \equiv Lw_k(z_l), \quad -r \leq k \leq s, \quad 1 \leq l \leq m,$$

and

$$(2.7) \quad [\mathcal{B}]_{k,l} \equiv Lv_k(z_l), \quad 1 \leq k \leq m-1, \quad 1 \leq l \leq m.$$

Thus the coefficients of the difference equation (2.1) can be efficiently computed by first

solving

$$(2.8a) \quad \mathcal{E}e = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{where } \mathcal{E} = \begin{pmatrix} \mathcal{B} \\ \gamma^T \end{pmatrix}$$

and $0 \in \mathbb{R}^m$ a zero vector, provided of course that \mathcal{E} is nonsingular. The coefficients d_i are then simply given by

$$(2.8b) \quad d = \mathcal{A}e.$$

To establish conditions for \mathcal{E} to be nonsingular we consider the operator L^0 defined by

$$L^0 y(x) \equiv y^{(n)}(x).$$

Also let the matrices \mathcal{A}^0 , \mathcal{B}^0 and \mathcal{E}^0 be defined as the corresponding quantities without superscript, but with L^0 replacing L . Using the definitions (2.3a, b) of the basis polynomials, together with the assumptions on the mesh S_h and the points z_i , it is then an easy matter to verify that $\det(\mathcal{E}) = \{1 + O(h)\} \det(\mathcal{E}^0)$. This proves the following.

THEOREM 2.1. *Let the coefficients of the operator L in (1.1) be continuous in an open interval containing $[0, 1]$. Then the coefficients d_i and e_i are well defined by (2.5) for all sufficiently small h provided that \mathcal{E}^0 is nonsingular.*

3. The order of consistency. Let $y(x)$ be the exact solution of (1.1) subject to appropriate initial or boundary conditions and assume that $y(x)$ is unique. The *local truncation error* of the difference approximation (2.1) is defined as

$$(3.1) \quad \tau_j \equiv \tau_j[y] \equiv \sum_{i=-r}^s d_i y(x_{j+i}) - \sum_{i=1}^m e_i L y(z_i).$$

If there is a positive constant c_1 and a positive integer ν , independent of h , such that $|\tau_j| \leq c_1 h^\nu$, with ν as high as possible, then the approximation is said to be consistent and the *order of consistency* is equal to ν . First the following basic result is shown.

THEOREM 3.1. *Let $y(x)$ be $r+s+m+n$ times continuously differentiable in some open interval containing $[0, 1]$. If \mathcal{E}^0 is nonsingular and $r+s \geq n$, then at least $n+1$ of the coefficients d_i are nonzero for all sufficiently small h and the order of consistency of (2.1) is greater than or equal to $r+s+m-n$.*

Proof. Let d_i^0 and e_i^0 be defined as the coefficients d_i and e_i , but with L^0 replacing L . Since $d_i = \{1 + O(h)\} d_i^0$, the first assertion of the theorem is true provided that at least $n+1$ of the coefficients d_i^0 are nonzero. So suppose on the contrary that $d_{i_k}^0 \neq 0$, $1 \leq k \leq k_1$, $k_1 < n+1$, and $d_i^0 = 0$ otherwise. Define $p(x) \equiv \{\prod_{k=1}^{k_1} (x - x_{j+l_k})\} q(x)$, where $q(x) \in \mathbb{P}_{n-k_1}$ is chosen such that $p(x)$ has degree n and leading coefficient $1/n!$. By assumption $n \leq r+s$, so certainly $n \leq r+s+m-1$. Hence by construction $\sum_{i=-r}^s d_i^0 p(x_{j+i}) = 0$ and $L^0 p(x) \equiv 1$. Therefore $\sum_{i=1}^m e_i^0 = 0$, which is a contradiction. To prove the second claim of the theorem we introduce points $\{t_i\}_{i=1}^{m-1}$ that satisfy $|t_i - x_j| \leq ch$, but that are otherwise arbitrary. Define

$$(3.2) \quad \mathcal{R}(x) \equiv \prod_{i=1}^{m-1} (x - t_i) \prod_{i=-r}^s (x - x_{j+i}),$$

and use the fact that $y(x) = \rho(x) + \mathcal{R}(x)\mathcal{D}(x)$ for some $\rho(x) \in \mathbb{P}_{r+s+m-1}$ and $\mathcal{D}(x) \in \mathcal{C}^n[x_j - \delta, x_j + \delta]$. Here $\delta > 0$ does not depend on h . (For example define $\rho(x)$ as a Lagrange or Hermite interpolant of $y(x)$.) Thus $\tau_j[y] = \tau_j[\rho] + \tau_j[\mathcal{R}\mathcal{D}] = \tau_j[\mathcal{R}\mathcal{D}] = -\sum e_i L\{\mathcal{R}(z_i)\mathcal{D}(z_i)\}$. It is not difficult to show that $e_i = O(1)$ since \mathcal{E}^0 is nonsingular, so that upon taking absolute values and estimating the individual terms the second conclusion of the theorem follows. \square

A necessary condition for \mathcal{E}^0 to be nonsingular is that $r + s \geq n$. For example the width $\omega \equiv r + s + 1$ of a difference approximation to a second order differential equation must be at least three. To prove the assertion suppose $r + s < n$. Then the n th derivative of each w_k vanishes identically, so that $d_i^0 = 0$ ($-r \leq i \leq s$). From the proof of Theorem 3.1 it then follows that \mathcal{E}^0 is singular. (If $\omega = n + 1$ then the difference equation is said to be compact.)

Some sufficient conditions for \mathcal{E}^0 to be nonsingular are given in the following theorem.

THEOREM 3.2. \mathcal{E}^0 is nonsingular if $m = 1$ and $r + s \geq n$, or if $r + s = n$.

Proof. The proof for the case $m = 1$ is immediate. If $r + s = n$, $m > 1$ and \mathcal{E}^0 singular then it follows from the definition of \mathcal{E}^0 that there are constants c_i ($1 \leq i \leq m$), not all zero, such that the polynomial $q(x) \equiv \sum_{i=1}^{m-1} c_i v_i(x) \in \mathbb{P}_{r+s+m-1}$ satisfies $q^{(n)}(z_i) = c_m$. Thus $\tilde{q}(x) \equiv q(x) - c_m x^n/n!$ satisfies $\tilde{q}^{(n)}(z_i) = 0$. But this is impossible since $\tilde{q}^{(n)}(x) \in \mathbb{P}_{m-1}$, because $r + s = n$. \square

It is well known (de Boor and Swartz (1973), Russell and Varah (1975)) that collocation procedures with certain piecewise polynomials for the numerical solution of boundary value problems have a higher order of accuracy if Gaussian points are used as collocation points. One expects that a similar statement applies to the choice of the points z_i in the finite difference approximations discussed in this paper.¹ In order to identify such special points, we note that from the proof of Theorem 3.1 it follows that the leading part of the truncation error is given by $\tau_j = -\sum_{i=1}^m e_i \mathcal{R}^{(n)}(z_i) \mathcal{D}(z_i)$, with $\mathcal{R}(x)$ as in (3.2). Thus the order of the truncation error increases by at least one if the z_i are chosen such that $\mathcal{R}^{(n)}(z_i) = 0$ ($1 \leq i \leq m$). This proves the following:

THEOREM 3.3. Let $y(x)$ be $r + s + m + n$ times continuously differentiable in some open interval containing $[0, 1]$. If \mathcal{E}^0 is nonsingular, $r + s \geq n$ and $\mathcal{R}^{(n)}(z_i) = 0$ ($1 \leq i \leq m$), then the order of consistency of the finite difference approximation (2.1) is greater than or equal to $r + s + m - n + 1$.

If $m = 1$ and the approximation compact, then $\mathcal{R}(x) = \prod_{i=-r}^s (x - x_{j+i})$. In this case there is only one possible choice of z_1 , for which a higher order of consistency is obtained. If $m = 1$ and $r + s > n$ then there are $r + s + 1 - n$ possible choices of this point. For the case $m > 1$, $r + s = n$ there is a $(m - 1)$ parameter family of points z_i for which the improved order is obtained. The parameters are the points t_i in the definition of $\mathcal{R}(x)$.

Theorem 3.3 gives a minimum value for the order of consistency. The question arises whether it is possible for the order to be higher than $r + s + m - n + 1$. For this to be the case for a general operator L it must at least be true for the particular case where $L = L^0$. Thus the special points z_i for which a higher order might be attained can be determined by requiring that

$$(3.3a) \quad \sum_{i=-r}^s d_i^0 p(x_{j+i}) = \sum_{i=1}^m e_i^0 L^0 p(z_i),$$

$$(3.3b) \quad \sum_{i=1}^m e_i^0 = 1,$$

for all polynomials $p(x)$ of degree as high as possible. Here not only the coefficients d_i^0 and e_i^0 are treated as unknown, but also the points z_i . Thus one expects that (3.3a, b) can be satisfied for all $p(x) \in \mathbb{P}_{r+s+2m-1}$. For the operator L^0 the order of consistency is then found to be greater than or equal to $r + s + 2m - n$. (If $r + s + 2m - 1$ is odd and the mesh uniform then another increase of one order can be obtained.) For example consider the case where $r = s = 1$ and $n = 2$. Thus we are considering compact approximations to a

¹For compact difference approximations certain superconvergence results have also been derived by Osborne (1967), (1974).

second order equation. Let the mesh be uniform. If $m = 3$ then equations (3.3a, b) yield a set of points z_i for which the order is 6 in case $L = L^0$. Numerical experiments show that the order for general L is also 6 with these z_i . If $m = 4$ then one finds a set of points z_i for which the order equals 8 when $L = L^0$. However for other L the order is observed to be only 6. For the two particular examples above, these phenomena are easily verified analytically, but we do not have a precise general statement.

4. Approximations to initial and boundary conditions. To define a complete difference scheme, approximations to initial or boundary conditions are required. Let such a condition be given by

$$(4.1) \quad By(0) \equiv y^{(k)}(0) + \sum_{i=0}^{k-1} b_i y^{(i)}(0) = b,$$

with $k < n$. Finite difference approximations to (4.1) are required to have the form

$$(4.2) \quad B_h u_0 \equiv \sum_{i=0}^s d_i u_i = b + \sum_{i=1}^m e_i f(z_i).$$

Here $f(x)$ is the inhomogeneous term of the differential equation (1.1). Note that by definition of the mesh S_h in § 1, x_0 need not coincide with $x = 0$. If $s = k$ then the approximation (4.2) is said to be *compact*. Proceeding as in § 2 we determine the coefficients d_i and e_i by requiring that (4.2) is satisfied for all $p(x) \in \mathbb{P}_{s+m}$, i.e.

$$(4.3) \quad \sum_{i=0}^s d_i p(x_i) = Bp(0) + \sum_{i=1}^m Lp(z_i).$$

For $p(x)$ we consecutively take the basis polynomials $w_k(x)$ ($0 \leq k \leq s$) and $v_k(x)$ ($1 \leq k \leq m$), which are defined in a fashion similar to those employed in § 2. More specifically we let the $w_k(x)$ be given by (2.3a) with $r = 0$ (and $j = 0$), while the remaining m basis polynomials are defined as

$$v_k(x) \equiv h^{s+k} x^{k-1} \prod_{l=0}^s (x - x_l), \quad 1 \leq k \leq m.$$

Since the coefficient of b in (4.2) is unity, no other normalization condition needs to be imposed. The equations (4.3) can now be written in matrix form as

$$(4.4) \quad \begin{pmatrix} I & -\mathcal{A} \\ O & -\mathcal{B} \end{pmatrix} \begin{pmatrix} d \\ e \end{pmatrix} = \begin{pmatrix} \beta_w \\ \beta_v \end{pmatrix},$$

with \mathcal{A} , d and e as in § 2 (with $r = 0$). \mathcal{B} is now a square matrix with entries

$$[\mathcal{B}]_{k,l} = Lv_k(z_l), \quad 1 \leq k \leq m, \quad 1 \leq l \leq m,$$

and we have introduced

$$\beta_w \equiv (Bw_0(0), \dots, Bw_s(0))^T \in \mathbb{R}^{s+1}$$

and

$$\beta_v \equiv (Bv_1(0), \dots, Bv_m(0))^T \in \mathbb{R}^m.$$

Also let \mathcal{B}^0 be defined as the corresponding quantity without superscript, but with L^0 replacing L . Then the following can be shown.

THEOREM 4.1. *Let the coefficients of L be continuous in some open interval containing $[0, 1]$. Then the coefficients d_i and e_i in (4.2) are uniquely defined by (4.4) for all sufficiently small h provided that \mathcal{B}^0 is nonsingular.*

Again due to the special choice of the basis, the solution of the system (4.4) is easily obtained by first solving for the e_i .

The truncation error of (4.2) is defined as

$$\tau_0 \equiv B_h y(0) - B y(0) - \sum_{i=1}^m e_i L y(z_i),$$

and the *order of consistency* is defined as in § 3. The proofs of the following two theorems closely follow those of Theorem 3.1 and Theorem 3.2 and are omitted. For details see Doedel (1976).

THEOREM 4.2. *Let $y(x)$ be $s + m + n + 1$ times continuously differentiable in some open interval containing $[0, 1]$. Assume that \mathcal{B}^0 is nonsingular. If $m = 0$ let $s \geq k$; otherwise if $m > 0$ let $s + m \geq n$. Then at least $k + 1$ of the coefficients d_i are nonzero for all sufficiently small h , and the order of consistency of the difference approximation (4.2) is greater than or equal to $s + m - k + 1$.*

Two important cases in which the main hypothesis of Theorem 4.2 is satisfied are given in the following theorem.

THEOREM 4.3. *\mathcal{B}^0 is nonsingular if $m = 0$ and $s \geq k$, or if $s = k$, $m > 0$ and $s + m \geq n$.*

As was the case for approximations to the differential equation, it is possible to identify points z_i for which the approximation to the boundary condition attains a higher order of consistency than predicted by Theorem 4.2. For this purpose define

$$\mathcal{R}_0(x) \equiv \prod_{i=1}^m (x - t_i) \prod_{i=0}^s (x - x_i),$$

for certain parameters t_i satisfying $|t_i| < c_1 h$. A minor modification of the proof of Theorem 3.3 then shows the following.

THEOREM 4.4. *Let $y(x)$ be $s + m + n + 1$ times continuously differentiable in some open interval containing $[0, 1]$. Assume that \mathcal{B}^0 is nonsingular. If $m = 0$ let $s \geq k$; otherwise if $m > 0$ let $s + m \geq n$. Also suppose that $\mathcal{R}_0^{(k)}(0) = 0$ and that $\mathcal{R}_0^{(n)}(z_i) = 0$ ($1 \leq i \leq m$). Then the order of consistency of the finite difference approximation (4.2) is greater than or equal to $s + m - k + 2$.*

5. Examples. In this section the results of previous sections are illustrated by means of a number of examples. First we consider difference approximations to the second order equation

$$(5.1) \quad y''(x) + a^1(x)y'(x) + a^0(x)y(x) = f(x).$$

Example 5.1. The choice $r = s = m = 1$, $z_1 = x_j$, and equal spacing yields the usual $O(h^2)$ central difference approximation. Use of a nonuniform mesh gives an $O(h)$ formula unless $z_1 = x_j + (h_{j+1} - h_j)/3$ which is the inflection point of $\mathcal{R}(x) = \prod_{i=-1}^1 (x - x_{j+i})$.

Example 5.2. A generalization of the well known three point Mehrstellenverfahren of Collatz (1966) is obtained with $r = s = 1$, $m = 3$, $z_1 = x_{j-1}$, $z_2 = x_j$ and $z_3 = x_{j+1}$. The order of this difference equation is 3, as predicted by Theorem 3.1. Explicit representation of the coefficients of this approximation is quite complicated. Moreover for numerical purposes it is much more efficient to have the computer solve the equations (2.5). If the mesh is uniform and $a^1(x) \equiv 0$ then the difference formula reduces to the usual form of the Mehrstellenverfahren. The order is then equal to 4.

Example 5.3. By Theorem 3.2, noncompact approximations to (5.1) with $m = 1$ are always consistent. For example if $r = s = 2$, $m = 1$, $z_1 = x_j$ and if the mesh is uniform then one obtains the usual 4th order centered five point approximation.

If $r + s > n$ and $m > 1$ then the resulting difference formula need not be consistent. As a first example consider the case in which $r = 1$, $s = 2$, $m = 1$ and the mesh uniform. Let $z_1 = x_{j-1} + \xi_1 h$ and $z_2 = x_{j-1} + \xi_2 h$. The determinant of \mathcal{E}^0 in Theorem 2.1 is equal to $c_1(\xi_2 - \xi_1)(\xi_2 + \xi_1 - 3)$ for some constant c_1 . By assumption the case $\xi_1 = \xi_2$ is excluded. Hence \mathcal{E}^0 is singular only if $\xi_1 + \xi_2 = 3$, i.e., when z_1 and z_2 are placed symmetrically about the midpoint. One cannot alleviate this problem by redefining the normalization. For example if one sets $e_2 = 1$ then the difference equation generated for the problem $y''(x) = f(x)$, with $\xi_1 = 1$ and $\xi_2 = 2$, is found to be $(-u_{j-1} + 3u_j - 3u_{j+1} + u_{j+2})/h^2 = f(x_{j+1}) - f(x_j)$. Except for a factor of h , this is consistent with $y'''(x) = f'(x)$.

As a second example take $r = 1$ and $s = 2$ as before, but let $m = 3$, $z_1 = x_{j+1/2} - \xi h$, $z_2 = x_{j+1/2}$ and $z_3 = x_{j+1/2} + \xi h$; $x_{j+1/2} \equiv \frac{1}{2}(x_j + x_{j+1})$. The determinant of \mathcal{E}^0 is now equal to $c_2 \xi^3 (4\xi^2 - 3)$. Hence, for distinct points \mathcal{E}^0 is singular if and only if $\xi = \frac{1}{2}\sqrt{3}$.

Example 5.4. In Table 5.1 we give three examples of special choices of the points z_i . The mesh is assumed to be uniform. For general ξ the order of the approximation is θ_1 . Theorem 3.3 predicts that the order is at least θ_2 , if $\xi = \xi_0$. The actual order observed in numerical computations is θ_3 . To determine ξ_0 in the 1st and 3rd example we have imposed the additional constraint that the points z_i be placed symmetrically. The particular choice of the z_i in the 2nd example has the advantage that the z_i of consecutive difference equations partly coincide, thereby limiting the necessary total number of function evaluations. Note that there are two possible values of ξ_0 in the last two examples.

TABLE 5.1

r	s	m	θ_1	$z_i (1 \leq i \leq m)$	ξ_0	θ_2	θ_3
1	1	2	2	$x_j \pm \xi h$	$1/\sqrt{6}$	3	4
1	1	4	4	$x_{j \pm 1/2} \pm \xi h$	$\{5/12 \pm \frac{1}{2}(23/45)^{1/2}\}^{1/2}$	5	6
2	2	2	4	$x_j \pm \xi h$	$\{1 \pm (11/15)^{1/2}\}^{1/2}$	5	6

Example 5.5. Two examples of approximations to the initial or boundary condition $By(0) \equiv y'(0) + b_0 y(0) = b$ that use the differential equation (5.1) are given in Table 5.2.

TABLE 5.2

s	m	θ_1	z_1	θ_2	θ_3
1	1	2	$h/3$	3	3
2	1	3	$h(9 \pm \sqrt{33})/12$	4	4

The mesh is uniform and $x_0 = 0$. The special choice of z_1 that yields the increased order θ_2 is obtained by applying Theorem 4.4. Since the 2nd approximation is not compact and since $m > 0$, this need not be consistent for all z_1 . In fact this approximation fails to be consistent if $z_1 = x_1$.

6. Stability for initial value problems. The purpose of this section is to indicate what form the well known stability analysis of difference methods for initial value problems takes, when applied to the class of difference methods studied in this paper. For background material we refer to Gear (1971) and Varah (1975). To keep the presentation simple the discussion is confined to first order equations. Thus the usual stability analysis leads us to consider the equation

$$(6.1) \quad \mathcal{L}(\lambda)y(x) \equiv y'(x) - \lambda y(x) = 0.$$

Difference approximations to (6.1) are constructed by the procedure of § 2. The mesh is assumed to be uniform. These approximations are written as

$$(6.2) \quad \mathcal{L}_h(\lambda)u_j \equiv \sum_{i=0}^{s_j} d_{j,i}(\lambda)u_{j+i} = 0, \quad j = 1, 2, \dots$$

The subscript j is henceforth omitted. The $d_i(\lambda)$ are determined by requiring that $\sum_{i=0}^s d_i(\lambda)p(x_{j+i}) - \sum_{i=0}^m e_i(\lambda)\mathcal{L}(\lambda)p(z_i) = 0$ for all $p(x) \in \mathbb{P}_{s+m}$, and we assume that \mathcal{E}^0 is nonsingular. A necessary condition for stability is that the roots η_i of the characteristic equation $c(\eta) \equiv h \sum_{i=0}^s d_i(\lambda)\eta^i = 0$ satisfy the root condition:

$$|\eta_i| \leq 1, \quad 1 \leq i \leq s,$$

and

$$|\eta_i| = 1 \quad \text{implies} \quad \eta_i \text{ is a simple root.}$$

It is not difficult to see that the characteristic polynomial has the form

$$(6.3) \quad c(\eta) = \sigma_0(\eta) + \sum_{i=1}^m (h\lambda)^i \sigma_i(\eta),$$

where $\sigma_0(\lambda) = h \sum_{i=0}^s d_i(0)\eta^i$ and where the remaining $\sigma_i(\eta)$ are also polynomials in η of degree at most s .

Stability properties of the "finite difference form" (6.3) have been extensively investigated for the case $m = 1$ by, for example, Dahlquist (1959), (1963), Gear (1971) and Varah (1975). The general case has been studied by Reimer (1968), Jeltsch (1975) and is still under further investigation. The motivation for considering such general finite difference forms with $m > 1$ is the fact that the stability analysis of many difference methods leads to studying such forms. Examples of these include methods based upon Padé rational approximations to the exponential (see Varga (1961)), Runge-Kutta methods, collocation methods and second derivative methods (Enright (1974).) Above we have indicated how these higher order finite difference forms arise in the stability analysis of the very general type of finite difference approximations considered in this paper. It is not our purpose here to contribute extensively to the investigations referred to above, but by means of some examples the effect that the choice of the points z_i has on the stability of the method will be illustrated.

First let $s = m = 1$ with $z_1 = x_j + \xi h$. So the approximation to (5.1) has the form $d_0(\lambda)u_j + d_1(\lambda)u_{j+1} = 0$, where the coefficients are found to be given by $d_0(\lambda) = -1/h - \lambda(1 - \xi)$ and $d_1(\lambda) = 1/h - \lambda\xi$. Therefore the characteristic polynomial is $c(\eta) \equiv h\{d_1(\lambda)\eta + d_0(\lambda)\} = \sigma_0(\eta) + h\lambda\sigma_1(\eta)$, where $\sigma_0(\eta) = \eta - 1$ and $\sigma_1(\eta) = -\xi\eta + \xi - 1$. The root of $\sigma_1(\eta) = 0$ is $\eta_1 = (\xi - 1)/\xi$, so that $|\eta_1| \leq 1$ if and only if $\xi \geq \frac{1}{2}$. Thus this method is stable at ∞ if $\xi \geq \frac{1}{2}$. It is easy to check that this difference approximation is in fact $A(\alpha)$ stable for $\xi \geq \frac{1}{2}$. Obviously the method is always strictly stable at zero.

Next let $s = 2$, $m = 1$ and $z_1 = x_j + \xi h$. Then $c(\eta) = \sigma_0(\eta) + h\lambda\sigma_1(\eta)$, where $\sigma_0(\eta) = (2\xi - 1)\eta^2 + 4(1 - \xi)\eta + 2\xi - 3$ and $\sigma_1(\eta) = \xi(1 - \xi)\eta^2 + 2\xi(\xi - 2)\eta - (\xi^2 - 3\xi + 2)$. Applying the analysis of § 3 yields that the order of this approximation is equal to 2 but increases to 3 if $\xi = 1 \pm \frac{1}{3}\sqrt{3}$. The roots of $\sigma_0(\eta) = 0$ are $\eta_1 = 1$ and $\eta_2 = (2\xi - 3)/(2\xi - 1)$. Now $|\eta_2| \leq 1$ if and only if $\xi \geq 1$. Thus this approximation is stable at zero if $\xi \geq 1$. In particular the third order method with $\xi = 1 + \frac{1}{3}\sqrt{3}$ and the well known Gear's method with $\xi = 2$ are strongly stable at zero. As for stability at ∞ , the roots of $\sigma_1(\eta) = 0$ are $\eta = [2\xi - \xi^2 \pm (2\xi - \xi^2)^{1/2}]/(\xi(1 - \xi))$, and some computation reveals that $|\eta_i| \leq 1$ provided that $\xi \geq 1 + \frac{1}{2}\sqrt{2}$. Thus the second order method of Gear is stable at ∞ , but the

third order formula corresponding to $\xi = 1 + \frac{1}{3}\sqrt{3}$ is not. A very interesting property of the method with $\xi = 1 + \frac{1}{2}\sqrt{2}$ is the fact that the region of stability coincides with the negative half plane.

7. Stability for boundary value problems. In this section the stability of finite difference schemes for boundary value problems is investigated. For this purpose the stability theory of Grigorieff (1970), or the quite similar but somewhat more general theory of Kreiss (1972), can be applied. The differential equation under consideration is given by (1.1) and the boundary conditions are

$$(7.1a) \quad B_k(0)y(0) \equiv \sum_{i=0}^{n_k(0)} b_{k,i}(0)y^{(i)}(0) = b_k(0), \quad 1 \leq k \leq n_0,$$

and

$$(7.1b) \quad B_k(1)y(1) \equiv \sum_{i=0}^{n_k(1)} b_{k,i}(1)y^{(i)}(1) = b_k(1), \quad 1 \leq k \leq n - n_0,$$

where

$$n_0 \geq 1, \quad n_k(0) < n \quad \text{and} \quad n_k(1) < n.$$

Throughout this section we assume that the mesh is uniform. A finite difference scheme consists of approximations to the differential equation (1.1) of the form

$$(7.2) \quad L_h u_j \equiv \sum_{i=-r}^s d_{j,i} u_{j+i} = \sum_{i=1}^m e_{j,i} f(z_{j,i}) \equiv \tilde{f}_j, \quad r \leq j \leq J-s,$$

together with discrete boundary conditions

$$(7.3a) \quad B_{h,k}(0)u_0 \equiv \sum_{i=0}^{s_k(0)} d_{k,i}(0)u_i = b_k(0) + \sum_{i=1}^{m_k(0)} e_{k,i}(0)f(z_{k,i}(0)) \equiv \tilde{b}_k(0), \quad 1 \leq k \leq n_0,$$

$$(7.3b) \quad B_{h,k}(1)u_J \equiv \sum_{i=-r_k(1)}^0 d_{k,i}(1)u_{J+i} = b_k(1) + \sum_{i=1}^{m_k(1)} e_{k,i}(1)f(z_{k,i}(1)) \\ \equiv \tilde{b}_k(1), \quad 1 \leq k \leq n - n_0.$$

In addition, if (7.2) is not compact, i.e., if $r+s > n$, then $r+s-n$ extra difference equations are required in order to match the number of equations and the number of unknowns. Although this is not necessary, we assume that these extra equations are also consistent with the differential equation and given by

$$(7.4) \quad L_h u_j \equiv \sum_{i=0}^{s_j} d_{j,i} u_i = \sum_{i=1}^{m_j} e_{j,i} f(z_{j,i}) \equiv \tilde{f}_j, \quad k_0 \leq j \leq r-1, \quad k_0 \geq 0,$$

$$(7.5) \quad L_h u_j \equiv \sum_{i=-r_j}^0 d_{j,i} u_{J+i} = \sum_{i=1}^{m_j} e_{j,i} f(z_{j,i}) \equiv \tilde{f}_j, \quad J-s+1 \leq j \leq J-n+k_0.$$

Let $u_h = (u_0, \dots, u_J)^T$. Then the equations (7.2) through (7.5) can be compactly written as

$$(7.6) \quad L_h u_h = f_h.$$

Here $f_h \in \mathbb{R}^{J+1}$ is the appropriate right hand side vector and L_h is a $(J+1) \times (J+1)$ matrix. Consistency of the equations (7.2), (7.4) and (7.5) is easily seen to imply that $|\tilde{f}_j - f(x_j)| \leq ch$. Also, let $e_h \in \mathbb{R}^{J+1}$ be the error vector, i.e., $e_h = (e_0, \dots, e_J)^T$ with $e_j = y_j - u_j$, and let τ_h be the vector of truncation errors. For $w_h \in \mathbb{R}^{J+1}$ let $\|w_h\| \equiv$

$\max_{0 \leq j \leq J} |w_j|$. If A_h is a $(J+1) \times (J+1)$ matrix then $\|A_h\|$ is the induced matrix norm, i.e., $\|A_h\| \equiv \max_{w_h \neq 0} \|A_h w_h\| / \|w_h\|$. The finite difference scheme (7.6) is said to be *consistent* if $\|\tau_h\| \leq C_1 h$ as $h \rightarrow 0$ and *stable* if L_h^{-1} exists for all small enough h and $\|L_h^{-1}\| \leq C_2$. Here C_1 and C_2 are constants that do not depend on h . The stability property is essentially determined by the difference approximation to the highest derivative. To be more specific, the notion of characteristic polynomial of (7.2) is needed. Let d_i^0 be as in Theorem 3.1, and define the *characteristic polynomial* $\tilde{c}(\eta)$ by

$$\tilde{c}(\eta) \equiv h^n \sum_{i=-r}^s d_i^0 \eta^{r+i}.$$

It is not difficult to show that the equation $\tilde{c}(\eta) = 0$ must have a root $\eta = 1$ of multiplicity n , because of consistency. Hence one can write $\tilde{c}(\eta) = (\eta - 1)^n c(\eta)$, where $c(\eta)$ is defined to be the *reduced characteristic polynomial* associated with the difference equation (7.2). Assume that $c(\eta)$ is explicitly given by

$$(7.7) \quad c(\eta) = \sum_{i=0}^N a_i \eta^i, \quad \text{with } N = r + s - n.$$

If the finite difference scheme (7.6) is not compact, then one also has $r + s - n$ reduced characteristic polynomials associated with the extra boundary conditions (7.4) and (7.5). It is assumed that these have the form

$$(7.8) \quad c_j(\eta) = \sum_{i=0}^{N_j} a_{j,i} \eta^i, \quad k_0 \leq j \leq r-1,$$

and

$$(7.9) \quad c_j(\eta) = \sum_{i=0}^{N_j} a_{j,i} \eta^i, \quad J-s+1 \leq j \leq J-n+k_0.$$

Finally, consider the homogeneous difference equation

$$(7.10) \quad \sum_{i=0}^N a_i v_{j+i} = 0, \quad 0 \leq j < \infty,$$

with boundary conditions

$$(7.11) \quad \sum_{i=0}^{N_j} a_{j,i} v_i = 0, \quad k_0 \leq j \leq r-1,$$

$$\sup_{0 \leq j \leq \infty} |v_j| \leq \text{const.}, \quad 0 \leq j < \infty,$$

and

$$(7.12) \quad \sum_{i=0}^N a_{N-i} v_{J-j-i} = 0, \quad 0 \leq j < \infty,$$

subject to

$$(7.13) \quad \sum_{i=0}^{N_j} a_{j,N_j-i} v_{J-i} = 0, \quad J-s+1 \leq j \leq J-n+k_0,$$

$$\sup_{0 \leq j < \infty} |v_{J-i}| \leq \text{const.}$$

Then one can state the following theorem due to Grigorieff (1970) and Kreiss (1972).

THEOREM 7.1. *Let the homogeneous problem given by (1.1) and (7.1a, b) only admit the trivial solution. Assume that the difference scheme (7.6) is consistent and that all roots η_i of the characteristic equation $c(\eta) = 0$ satisfy $|\eta_i| \neq 1$. Further, if the difference scheme is not compact then also assume that the difference equations (7.10) with boundary conditions (7.11) and the difference equations (7.12) with boundary conditions (7.13) only have the trivial solution. Then (7.6) has a unique solution for all sufficiently small h and the difference scheme is stable.*

We first investigate the stability properties of the approximation (7.2). If (7.2) is compact (and consistent) then consistency of the boundary conditions (7.3a, b) is sufficient to guarantee stability. If (7.2) is not compact then the reduced characteristic polynomial $c(\eta)$ said to be *symmetric* if $c(\eta) = \eta^{r+s-n} c(1/\eta)$ and *strictly diagonally dominant* if the degree of $c(\eta)$ is even and if $|a_l| > \sum_{i=0, i \neq l}^{r+s-n} |a_i|$. (Here $l \equiv (r+s-n)/2$.) To motivate this last definition, consider the general form of a five point formula that is consistent with the second derivative. This approximation can be written as $K_h D_h^2 u_j$ where $D_h^2 u_j = (u_{j+1} - 2u_j + u_{j-1})/h^2$ and $K_h w_j = a_0 w_{j-1} + a_1 w_j + a_2 w_{j+1}$. The reduced characteristic polynomial is $c(\eta) = a_0 + a_1 \eta + a_2 \eta^2$ and $c(\eta)$ is diagonally dominant if and only if K_h is diagonally dominant. Except for approximations at the boundary, one normally constructs the difference approximation (7.2) such that $c(\eta)$ is symmetric.

LEMMA 7.1. *Assume that $c(t)$ is symmetric. If the degree of $c(\eta)$ is odd then $c(-1) = 0$. If the degree of $c(\eta)$ is even and if $c(\eta)$ is strictly diagonally dominant with positive coefficients, then there are no roots of $c(\eta) = 0$ on the unit circle.*

Proof. The proof of the first assertion is immediate. If the degree of $c(\eta)$ equals $2N_1$, for some integer N_1 , then

$$\begin{aligned} |c(e^{i\theta})| &= \left| \sum_{k=0}^{2N_1} a_k e^{ik\theta} \right| = \left| \left\{ a_{N_1} + \sum_{k=0}^{N_1-1} a_k (e^{ik\theta} + e^{-ik\theta}) \right\} e^{iN_1\theta} \right| \\ &= \left| a_{N_1} + 2 \sum_{k=0}^{N_1-1} a_k \cos(k\theta) \right| \geq a_{N_1} - 2 \left| \sum_{k=0}^{N_1-1} a_k \cos(k\theta) \right| \\ &\geq a_{N_1} - 2 \sum_{k=0}^{N_1-1} a_k > 0. \quad \square \end{aligned}$$

Example 7.1. In the special case where $c(\eta) = a_0 \eta^2 + (1 - 2a_0)\eta + a_0$ the assumptions of the lemma hold if $0 \leq a_0 < \frac{1}{4}$. A simple computation shows that in fact there are no roots on the unit circle iff $-\infty < a_0 < \frac{1}{4}$. This shows that the assumptions in the lemma are not strictly necessary, although perhaps desirable.

Example 7.2. Let $r = s = m = n = 2$, with $z_1 = x_j - \xi h$ and $z_2 = x_j + \xi h$. The reduced characteristic polynomial is given in Example 7.1 with $a_0 = (6\xi^2 - 1)/12$. Thus there are no roots of $c(\eta) = 0$ on the unit circle if and only if $0 < \xi < \frac{1}{3}\sqrt{6}$. If $\xi = \xi_0 \equiv \{1 \pm (11/15)^{1/2}\}^{1/2}$ then the order of the difference formula is 6. (See Example 5.4.) Therefore the stability theory guarantees stability only for the smaller value of ξ_0 . This does not imply that a finite difference scheme based upon the larger value of ξ_0 is necessarily unstable. That such a scheme may be stable is supported by numerical experiments in the next section. (See Table 8.1, experiment 14.) The smaller value of ξ_0 appears to give a better error constant however.

Now consider the extra boundary conditions (7.4) and (7.5). Again, assume that the reduced characteristic polynomial $c(\eta)$ of (7.2) is symmetric and that the degree of $c(\eta)$ is even and equal to $2N_1$. If in addition $c(\eta)$ is strictly diagonally dominant with positive coefficients then the characteristic equation $c(\eta) = 0$ has exactly N_1 roots inside

the unit circle and N_1 roots outside the unit circle. A necessary condition for the difference equations (7.10), (7.11) and (7.12), (7.13) to admit the zero solution only is then that the number of extra boundary conditions at $x = 0$ is the same as the number at $x = 1$ and equal to $N_1 = (r + s - n)/2$. It is also reasonable to assume now that the reduced characteristic polynomials of the extra equations at $x = 0$ and $x = 1$ are related by

$$(7.14) \quad c_{J+r-s-j}(\eta) = \eta^{N_1} c_j\left(\frac{1}{\eta}\right), \quad (r + s - n)/2 = k_0 \leq j \leq r - 1,$$

i.e., the conditions at $x = 1$ are the "reflection" of those at $x = 0$. For stability it is then sufficient to show that the difference equations (7.10) subject to (7.11) have the zero solution only. For this purpose define polynomials $p_j(\eta)$ by

$$(7.15) \quad p_j(\eta) \equiv \sum_{i=0}^N a_i \eta^{j+i}, \quad j \geq 0.$$

Here $N = r + s - n$ and the coefficients a_i are the same as those of the reduced characteristic polynomial $c(\eta)$ in (7.6). If (7.10) subject to (7.11) admits a nontrivial solution then it is easily seen that the polynomials $c_j(\eta)$ ($(r + s - n)/2 = k_0 \leq j \leq r - 1$) and $p_j(\eta)$ ($0 \leq j \leq \max N_j - N$) are linearly dependent. Hence we have shown the following:

THEOREM 7.2. *Let the homogeneous problem corresponding to (1.1) and (7.1a, b) only have the trivial solution and let the difference scheme (7.6) be consistent. Assume that $c(\eta)$ is symmetric. Also suppose that the degree of $c(\eta)$ is even and that $c(\eta)$ is strictly diagonally dominant with positive coefficients. Let the reduced characteristic polynomials of the extra boundary conditions be related as in (7.14). If the polynomials $c_j(\eta)$ and $p_j(\eta)$ defined by (7.8) and (7.15) respectively are linearly independent then the difference scheme is stable. Hence there exists a constant K independent of h such that $\|e_h\| \leq K \|\tau_h\|$.*

Example 7.3. Let $c(\eta) = a_0 \eta^2 + (1 - 2a_0)\eta + a_0$ with $-\infty < a_0 < \frac{1}{4}$. If the degree of the reduced characteristic polynomial $c_1(\eta)$ of the extra boundary condition at $x = 0$ is also equal to two then the difference scheme is stable if $c_1(\eta) \neq c(\eta)$. If the degree of $c_1(\eta)$ is three then stability is guaranteed if there are no constants α_1 and α_2 such that $c_1(\eta) = \alpha_1(a_0 \eta^3 + (1 - 2a_0)\eta^2 + a_0 \eta) + \alpha_2(a_0 \eta^2 + (1 - 2a_0)\eta + a_0)$.

8. Numerical examples. The main purpose of the numerical examples given in this section is to check the correctness of statements in previous sections. They also give some indication as to what the relative accuracy of various discretizations is. All computations were carried out on an IBM 370/168, using double precision arithmetic. No attempt was made to optimize the efficiency of the computations, so that there will be no conclusions about the relative merit of various finite difference schemes.

Example 8.1. Let the differential equation be given by

$$Ly(x) \equiv y'' + y' - 2y = 2(1 - 6x)e^x$$

with boundary conditions

$$y(0) = y(1) = 0.$$

The solution of this problem is

$$y(x) = 2x(1 - x)e^x.$$

The results of some numerical computations are given in Table 8.1. In this table r , s and m are as in the finite difference approximation (2.1). So the width of the approximation

equals $\omega \equiv r + s + 1$ and m denotes the number of points z_i . The letters A, B, C, D and E in the columns headed by "c" are a code indicating the location of the z_i , viz.

- A: The points z_i ($1 \leq i \leq m$) are optimal i.e., the order of consistency is as high as possible for the particular values of r , s and m considered.
- B: The points z_i ($1 \leq i \leq m$) are optimal under the restrictions that each subinterval contain the same number of points, and that these points are placed symmetrically with respect to the subinterval. (See the second case discussed in Example 5.4.)
- C: The points z_i coincide with the meshpoints and $m = r + s + 1$.
- D: $m = 1$ and $z_1 = x_j$.
- E: The placement of the points is not optimal, but they are placed symmetrically in the interval $[x_{j-n}, x_{j+s}]$.

The column headed by "o" gives the order of consistency of the finite difference approximation as predicted by theorems in preceding sections. Columns 7 through 11 define the finite difference equations (2.1) for $r \leq j \leq J - s$. If the width of this approximation is equal to 5 then a special finite difference equation must be defined for $j = 1$. This is done in columns 2–6. (The special equation necessary for $j = J - 1$ is assumed to be the "reflection" of the one for $j = 1$.) The mesh is taken uniform in this example, so that $h_j = h = 1/J$. For a number of values of J the observed maximum error, i.e., $\max |u_j - y(x_j)|$, is given. The notation 0.438^{-1} means $0.438 \cdot 10^{-1}$ etc. In the final column, headed by "a", the observed asymptotic order of accuracy of the given finite difference scheme is listed. If it is not clear from the numerical results what this order is then the expected order is given between brackets.

Most of the results that appear in the table are self explanatory. The first seven experiments involve compact difference approximations. For experiments 5 and 6 the points z_i are given by $z_i = x_{j \pm 1/2} \pm \xi h$. There are two values of ξ for which the optimal order of consistency is reached. (See the second case discussed in Example 5.4). These values are

$$\xi = \sqrt{\frac{5}{12} - \frac{1}{2}\sqrt{\frac{23}{45}}},$$

used in Experiment 5, and

$$\xi = \sqrt{\frac{5}{12} + \frac{1}{2}\sqrt{\frac{23}{45}}},$$

used in 6.

Experiments 8–12 show the effect that various choices of the extra boundary conditions have on the overall accuracy. Note that even if the order of consistency of the extra finite difference equations is only equal to two, then the order of accuracy of the scheme remains four. This phenomenon is also explained in the paper of Kreiss (1972). (See also Bramble and Hubbard (1964) and Shoosmith (1975).) The actual accuracy however is usually seriously effected.

In experiments 13 and 14 the points z_i of the main finite difference equations are $z_1 = x_j - \xi h$ and $z_2 = x_j + \xi h$. Again, as has been mentioned previously in Example 5.4, there are two values of ξ for which the order of consistency becomes six. These are

$$\xi = \sqrt{1 - \sqrt{\frac{11}{15}}}$$

used in 13 and

$$\xi = \sqrt{1 + \sqrt{\frac{11}{15}}}$$

used in 14.

TABLE 8.1

#		r	s	m	c	o	$J=4$	$J=8$	$J=16$	$J=32$	$J=64$	a
1		1	1	2	D	2	$.438^{-1}$	$.111^{-1}$	$.281^{-2}$	$.704^{-3}$	$.176^{-3}$	2
2		1	1	2	B	2	$.111^{-1}$	$.279^{-2}$	$.700^{-3}$	$.175^{-3}$	$.439^{-4}$	2
3		1	1	2	A	3	$.830^{-4}$	$.534^{-5}$	$.337^{-6}$	$.211^{-7}$	$.132^{-8}$	4
4		1	1	3	C	4	$.516^{-3}$	$.325^{-4}$	$.203^{-5}$	$.128^{-6}$	$.798^{-8}$	4
5		1	1	4	B	5	$.213^{-6}$	$.337^{-8}$	$.528^{-10}$	$.817^{-12}$	**	6
6		1	1	4	B	5	$.704^{-6}$	$.106^{-7}$	$.165^{-9}$	$.257^{-11}$	**	6
7		1	1	5	E	6	$.156^{-6}$	$.249^{-8}$	$.390^{-10}$	$.628^{-12}$	**	6
8	1	1	1	D	2	2	$.291^{-1}$	$.265^{-2}$	$.204^{-3}$	$.147^{-4}$	$.987^{-6}$	4
9	1	3	1	D	3	2	$.440^{-2}$	$.233^{-3}$	$.897^{-5}$	$.310^{-6}$	$.160^{-7}$	(4)
10	1	4	1	D	4	2		$.369^{-4}$	$.458^{-5}$	$.326^{-6}$	$.207^{-7}$	4
11	1	1	2	A	3	2	$.533^{-3}$	$.719^{-4}$	$.518^{-5}$	$.336^{-6}$	$.212^{-7}$	4
12	1	1	3	C	4	2	$.884^{-3}$	$.781^{-4}$	$.528^{-5}$	$.338^{-6}$	$.212^{-7}$	4
13	1	1	5	E	6	2	$.571^{-7}$	$.185^{-8}$	$.367^{-10}$	$.601^{-12}$	**	6
14	1	1	5	E	6	2	$.260^{-4}$	$.745^{-6}$	$.112^{-7}$	$.160^{-9}$	$.215^{-11}$	(6)

** Contaminated by roundoff.

Example 8.2. Consider the equation

$$y'' + xy' - (1+x)y = -(2+x)e^x,$$

with boundary conditions

$$y'(0) = y(1) = 0.$$

The solution to this problem is $y(x) = (1-x)e^x$.

In this example only compact approximations to the differential equations are considered. Numerical test calculations are performed with various difference approximations to the boundary condition $y'(0) = 0$. Results appear in Table 8.2. The notation used is the same as in the previous example. The approximation to the boundary condition is defined in columns 2–6, while the finite difference approximation to the differential equation is defined in the next five columns.

TABLE 8.2

#	r	s	m	c	o	r	s	m	c	o	$J=4$	$J=8$	$J=16$	$J=32$	$J=64$	a
1	0	1	0	—	1	1	1	1	D	2	$.117^0$	$.494^{-1}$	$.226^{-1}$	$.108^{-1}$	$.526^{-2}$	1
2	0	2	0	—	2	1	1	1	D	2	$.240^{-1}$	$.492^{-2}$	$.110^{-2}$	$.261^{-3}$	$.633^{-4}$	2
3	0	1	1	*	2	1	1	1	D	2	$.250^{-1}$	$.631^{-2}$	$.158^{-2}$	$.396^{-3}$	$.989^{-4}$	2
4	0	1	1	**	2	1	1	1	D	2	$.240^{-1}$	$.492^{-2}$	$.110^{-2}$	$.261^{-3}$	$.633^{-4}$	2
5	0	1	0	—	1	1	1	2	A	3	$.106^0$	$.466^{-1}$	$.219^{-1}$	$.106^{-1}$	$.521^{-2}$	1
6	0	2	0	—	2	1	1	2	A	3	$.362^{-1}$	$.786^{-2}$	$.183^{-2}$	$.443^{-3}$	$.109^{-3}$	2
7	0	1	1	*	2	1	1	2	A	3	$.152^{-1}$	$.359^{-2}$	$.876^{-3}$	$.216^{-3}$	$.538^{-4}$	2
8	0	1	1	**	2	1	1	2	A	3	$.338^{-1}$	$.764^{-2}$	$.181^{-2}$	$.440^{-3}$	$.108^{-3}$	2
9	0	1	1	A	3	1	1	2	A	3	$.531^{-3}$	$.602^{-4}$	$.711^{-5}$	$.862^{-6}$	$.106^{-6}$	3
10	0	1	2	C	3	1	1	2	A	3	$.146^{-2}$	$.171^{-3}$	$.207^{-4}$	$.254^{-5}$	$.316^{-6}$	3
11	0	4	0	—	4	1	1	2	A	3	$.344^{-2}$	$.164^{-3}$	$.898^{-5}$	$.526^{-6}$	$.317^{-7}$	4
12	0	1	3	E	4	1	1	2	A	3	$.482^{-4}$	$.324^{-5}$	$.208^{-6}$	$.132^{-7}$	$.828^{-9}$	4
13	0	2	1	A	4	1	1	2	A	3	$.165^{-4}$	$.108^{-5}$	$.730^{-7}$	$.472^{-8}$	$.300^{-9}$	4
14	0	2	1	A	4	1	1	2	A	3	$.563^{-3}$	$.249^{-4}$	$.129^{-5}$	$.734^{-7}$	$.437^{-8}$	4

* $z_1 = x_0$ ** $z_1 = x_1$

In experiment 9 the point z_1 for the boundary condition is $z_1 = x_0 + h/3$. For this value of z_1 the order of consistency is three. (See Example 5.5.) In experiments 13 and 14 this point is $z_1 = x_0 + \xi h$. In 13 the value of ξ is $(9 - \sqrt{33})/12$ and in 14 this value is $(9 + \sqrt{33})/12$. For these points the order of consistency is equal to four rather than three. Note that the order of accuracy is not greater than the order of consistency of the discrete boundary condition. This differs from observations made about the extra boundary conditions in Example 8.1.

Acknowledgments. The author wishes to thank Professor J. M. Varah at the University of British Columbia for guiding the Ph.D. thesis upon which this paper is based. He is also thankful to the first referee for detailed suggestions concerning presentation and to the second referee for making the author aware of previous work by M. R. Osborne.

REFERENCES

- G. BIRKHOFF AND S. GULATI (1974), *Optimal few point discretizations of linear source problems*, this Journal, 11, pp. 700–728.
- J. H. BRAMBLE AND B. E. HUBBARD (1964), *On a finite difference analogue of an elliptic boundary value problem which is neither diagonally dominant nor of the nonnegative type*, J. Math. and Phys., 43, pp. 117–132.
- L. COLLATZ (1966), *The Numerical Treatment of Differential Equations*, Springer-Verlag, New York.
- G. DAHLQUIST (1959), *Stability and error boundary in the numerical integration of ordinary differential equations*, Kungl. Tekn. Högsk. Stockholm, no. 130.
- (1963), *A special stability problem for linear multistep methods*, BIT, 3, pp. 27–43.
- C. DE BOOR AND B. SWARTZ (1973), *Collocation at Gaussian points*, this Journal, 10, pp. 582–606.
- E. J. DOEDEL (1976), *Difference methods for ordinary differential equations with applications to parabolic equations*, Thesis, University of British Columbia.
- (1977), *Finite difference methods for nonlinear two point boundary value problems*, manuscript.
- W. H. ENRIGHT (1974), *Second derivative multistep methods for stiff ordinary differential equations*, this Journal, 11, pp. 321–331.
- C. W. GEAR (1971), *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ.
- R. D. GRIGORIEFF (1970), *Die Konvergenz des Rand und Eigenwert problems linearer gewöhnlicher Differenzengleichungen*, Numer. Math., pp. 15–48.
- R. JELTSCH (1975), *Multistep multidervative methods and Hermite–Birkhoff interpolation*, manuscript.
- H. B. KELLER (1968), *Numerical Methods for Two Point Boundary Value Problems*, Blaisdell, London.
- (1976), *Numerical Solution of Two Point Boundary Value Problems*, Regional Conference Series in Applied Mathematics, no. 24, Society for Industrial and Applied Mathematics, Philadelphia.
- H. B. KELLER AND V. PEREYRA (1976), *Difference methods and deferred corrections for ordinary boundary value problems*, manuscript.
- H.-O. KREISS (1972), *Difference approximations for boundary and eigenvalue problems for ordinary differential equations*, Math. Comput., 26, pp. 605–624.
- P. LANCASTER (1969), *Theory of Matrices*, Academic Press, New York.
- R. E. LYNCH AND J. R. RICE (1975), *The HODIE method*, Rep. CSD-TR 170, Dept. of Computer Science, Purdue Univ., W. Lafayette, IN.
- (1976), *The HODIE method for ordinary differential equations*, Rep. CSD-TR 188, Dept. of Computer Science, Purdue Univ., W. Lafayette, IN.
- M. R. OSBORNE (1964), *A method for finite difference approximation to ordinary differential equations*, Computer J., 7, pp. 58–65.
- (1967), *Minimising truncation error in finite difference approximations to ordinary differential equations*, Math. Comput., 11, pp. 133–145.
- (1974), *Collocation, difference equations, and stitched function representations*, Proc. Dublin Conf. in Numerical Analysis.
- M. REIMER (1968), *Finite difference forms containing derivatives of higher order*, this Journal, 5, pp. 725–738.
- R. D. RUSSELL AND J. M. VARAH (1975), *A comparison of global methods for linear two point boundary value problems*, Math. Comput., 29, pp. 1–13.

- J. N. SHOOSMITH (1975), *A higher order finite difference method for the solution of two-point boundary value problems on a uniform mesh*, Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, A. K. Aziz, ed., Academic Press, New York, pp. 355–369.
- B. K. SWARTZ (1974), *The construction and comparison of finite difference analogs of some finite element schemes*, report, Los Alamos Scientific Laboratory, Los Alamos, NM.
- J. M. VARAH (1975), *Stiffly stable linear multistep methods for extended order*, Tech. Rep., Dept. of Comp. Sci., University of British Columbia.
- R. S. VARGA (1961), *On higher order stable implicit methods for solving parabolic partial differential equations*, J. Math. and Phys., 40, pp. 220–231.